greatlearning
Power Ahead

Business Presentation

Presenter : Esteban Ordenes

Contents

1. Data Definition and perform an Exploratory Data Analysis
2. Illustration of the insights based on EDA
3. Model building - Logistic Regression
4. Model building - Decision Tree Regression
5. Model Analysis
6. Actionable Insights & Recommendations

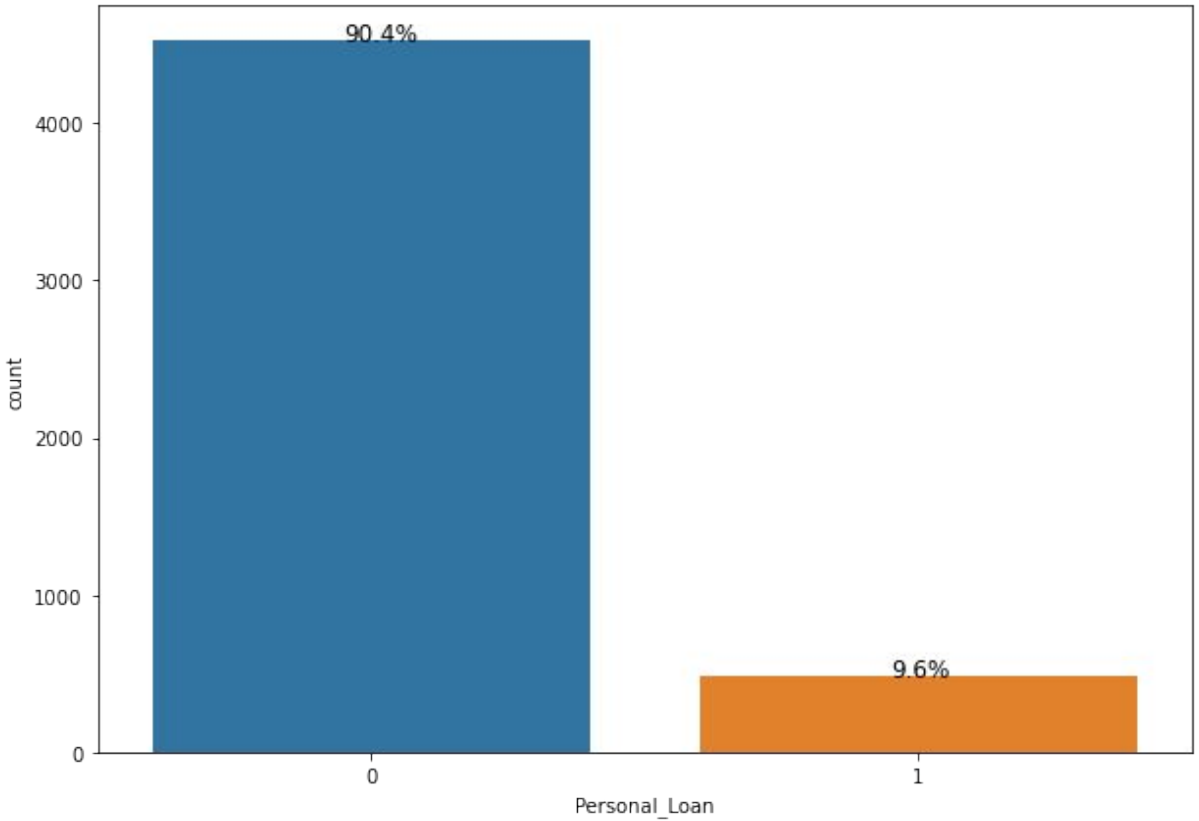
Business Problem Overview and Solution Approach

- Core business idea
  - A campaign that the bank ran last year for liability customers showed a healthy conversion rate of over 9% success. This has encouraged the retail marketing department to devise campaigns with better target marketing to increase the success ratio.
- Problem to tackle
  - Build a model that will help the marketing department to identify the potential customers who have a higher probability of purchasing the loan.

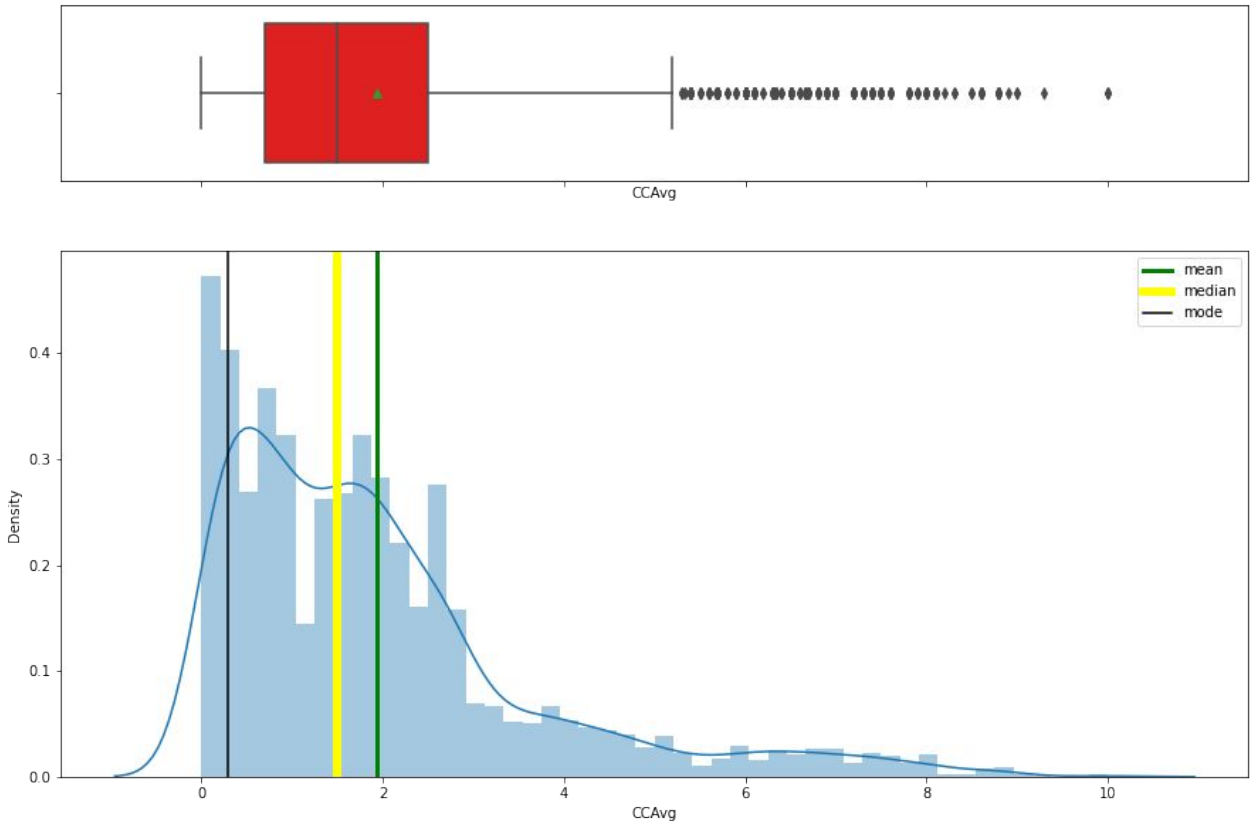
Data Overview

- **Data Dictionary**
- ID: Customer ID
- Age: Customer's age in completed years
- Experience: #years of professional experience
- Income: Annual income of the customer (in thousand dollars)
- ZIP Code: Home Address ZIP code.
- Family: the Family size of the customer
- CCAvg: Average spending on credit cards per month (in thousand dollars)
- Education: Education Level. 1: Undergrad; 2: Graduate;3: Advanced/Professional
- Mortgage: Value of house mortgage if any. (in thousand dollars)
- Personal_Loan: Did this customer accept the personal loan offered in the last campaign?
- Securities_Account: Does the customer have securities account with the bank?
- CD_Account: Does the customer have a certificate of deposit (CD) account with the bank?
- Online: Do customers use internet banking facilities?
- CreditCard: Does the customer use a credit card issued by any other Bank (excluding All life Bank)

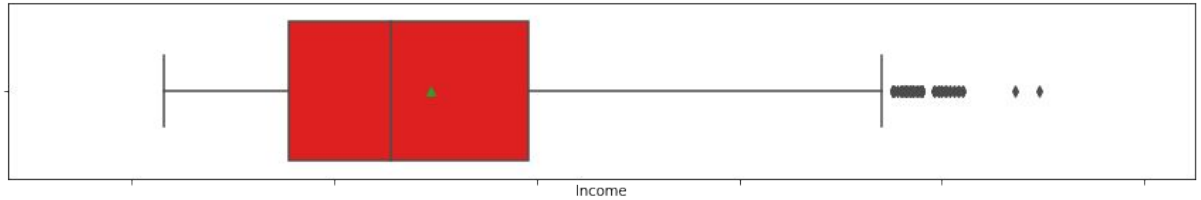
EDA

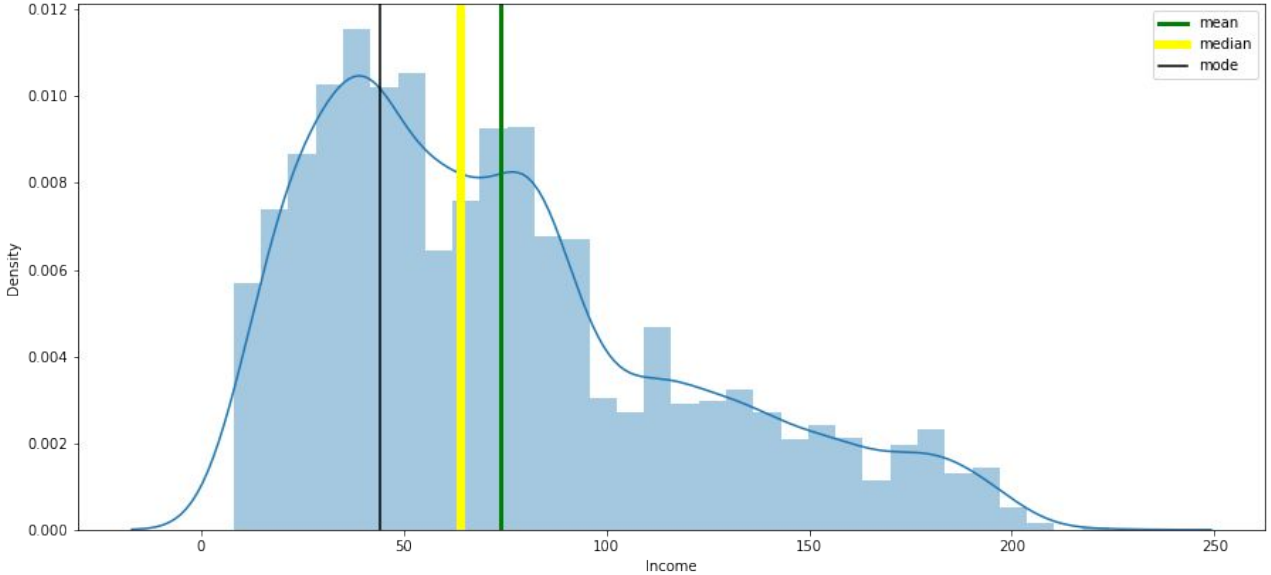


EDA

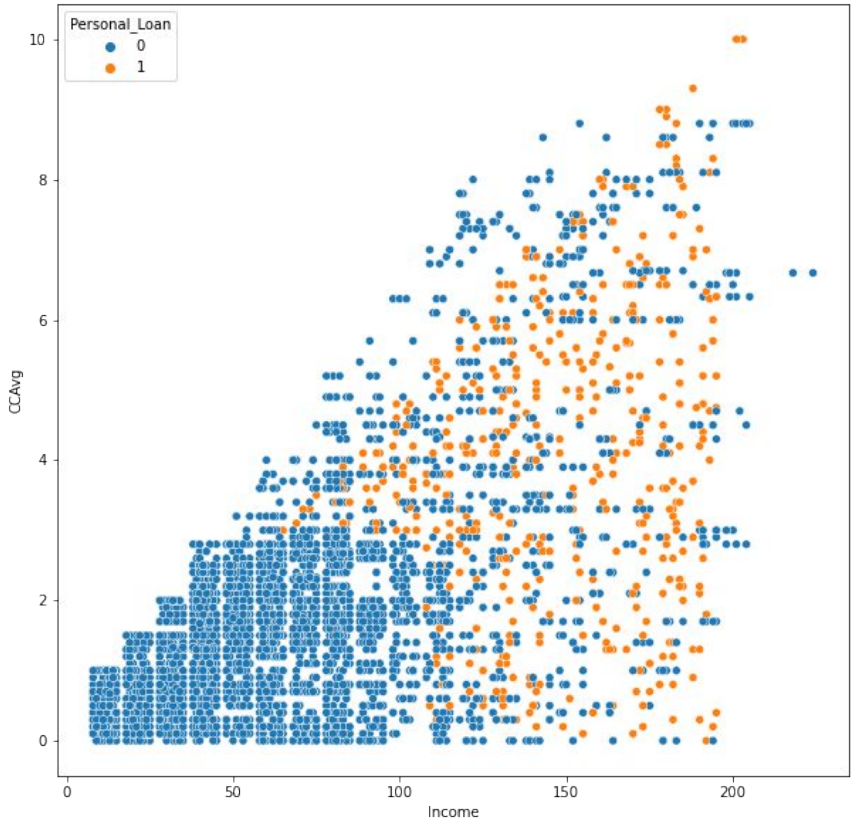


EDA

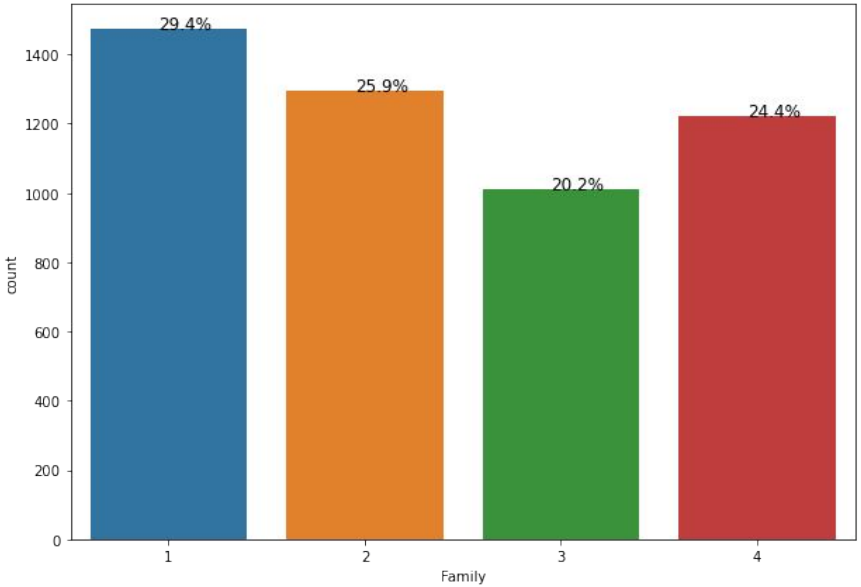


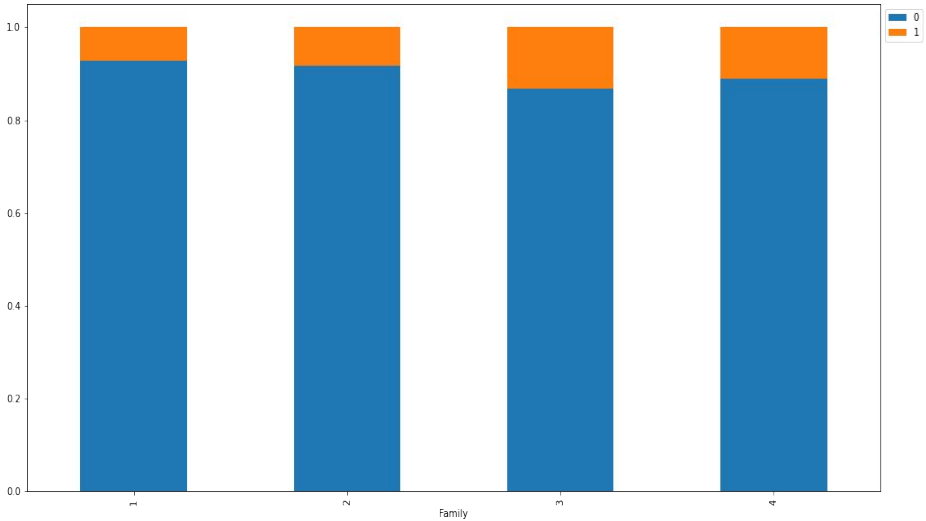


EDA

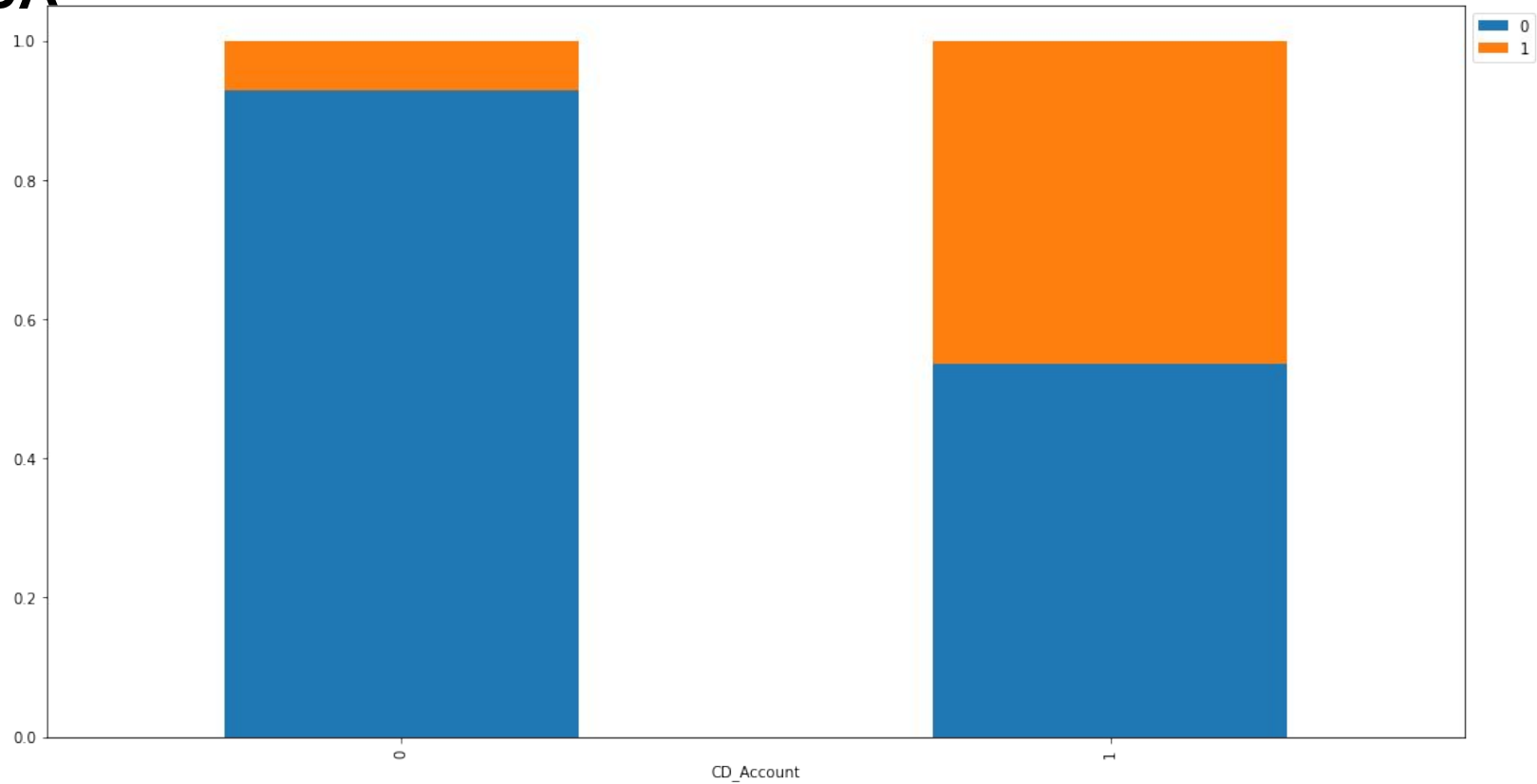


EDA



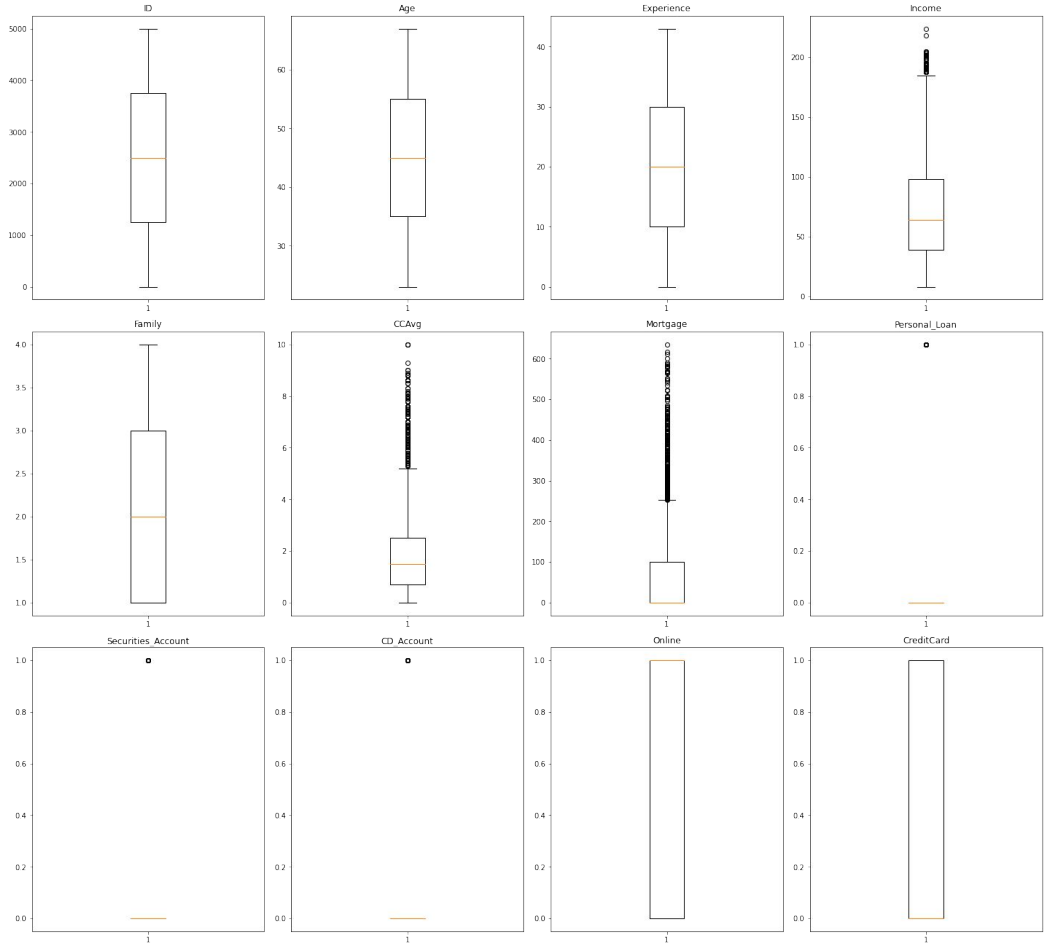


EDA



EDA

EDA



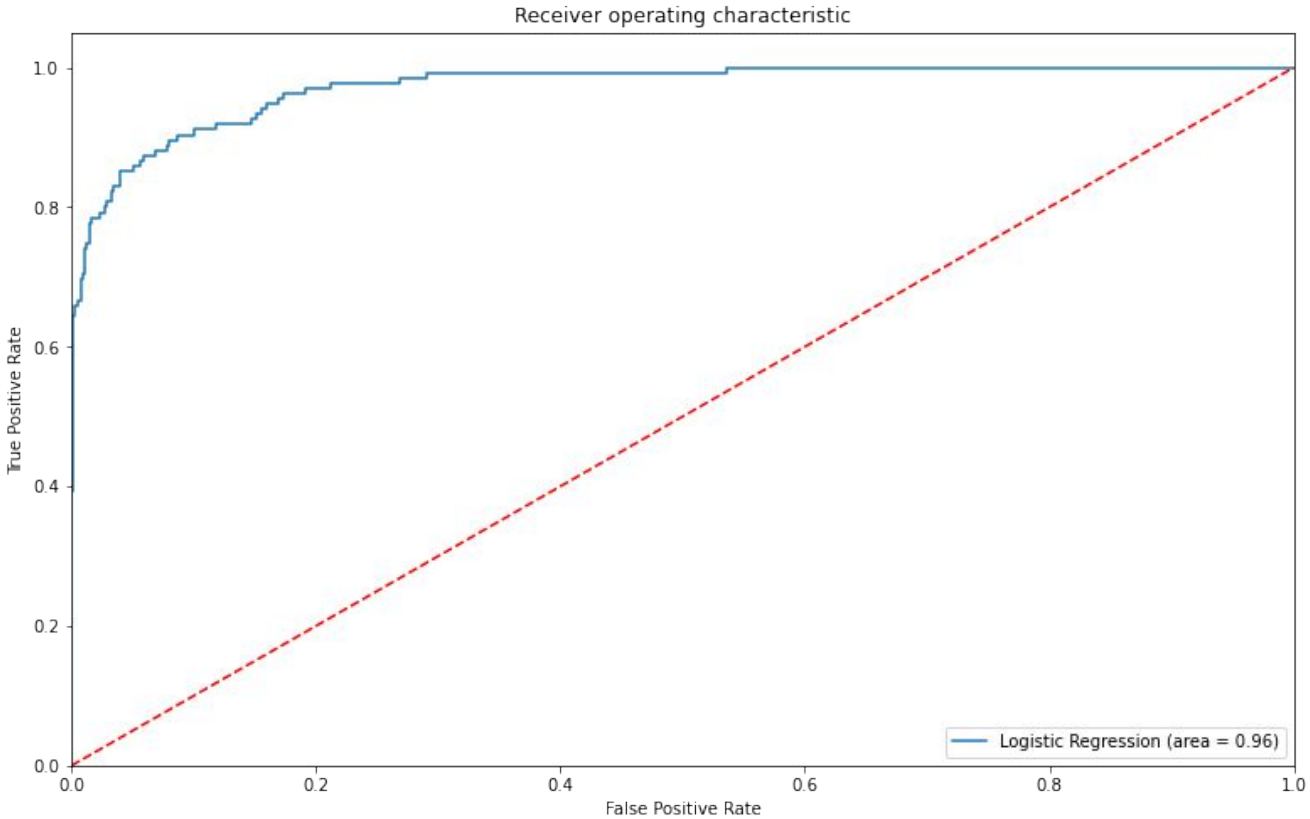
Logistic Regression Model Performance Summary

Logit Regression Results

Dep. Variable:	Personal_Loan	No. Observations:	3500
Model:	Logit	Df Residuals:	3482
Method:	MLE	Df Model:	17
Date:	Sat, 08 May 2021	Pseudo R-squ.:	0.6292
Time:	04:31:00	Log-Likelihood:	-412.06
converged:	True	LL-Null:	-1111.2
Covariance Type:	nonrobust	LLR p-value:	3.645e-287

	coef	std err	z	P>|z|	[0.025	0.975]
const	-13.1444	0.739	-17.780	0.000	-14.593	-11.695
Experience	0.0066	0.008	0.808	0.419	-0.009	0.023
Income	0.0583	0.004	16.401	0.000	0.051	0.065
Family	0.5781	0.090	6.402	0.000	0.401	0.755
CCAvg	0.4374	0.069	6.381	0.000	0.303	0.572
Mortgage	0.0010	0.001	1.055	0.292	-0.001	0.003
Securities_Account	-0.8714	0.360	-2.420	0.016	-1.577	-0.166
CD_Account	3.7311	0.416	8.972	0.000	2.916	4.546
Online	-0.8127	0.200	-4.072	0.000	-1.204	-0.422
CreditCard	-1.1725	0.257	-4.568	0.000	-1.676	-0.669
Education_2	3.9897	0.320	12.474	0.000	3.363	4.617
Education_3	3.9409	0.309	12.754	0.000	3.335	4.546

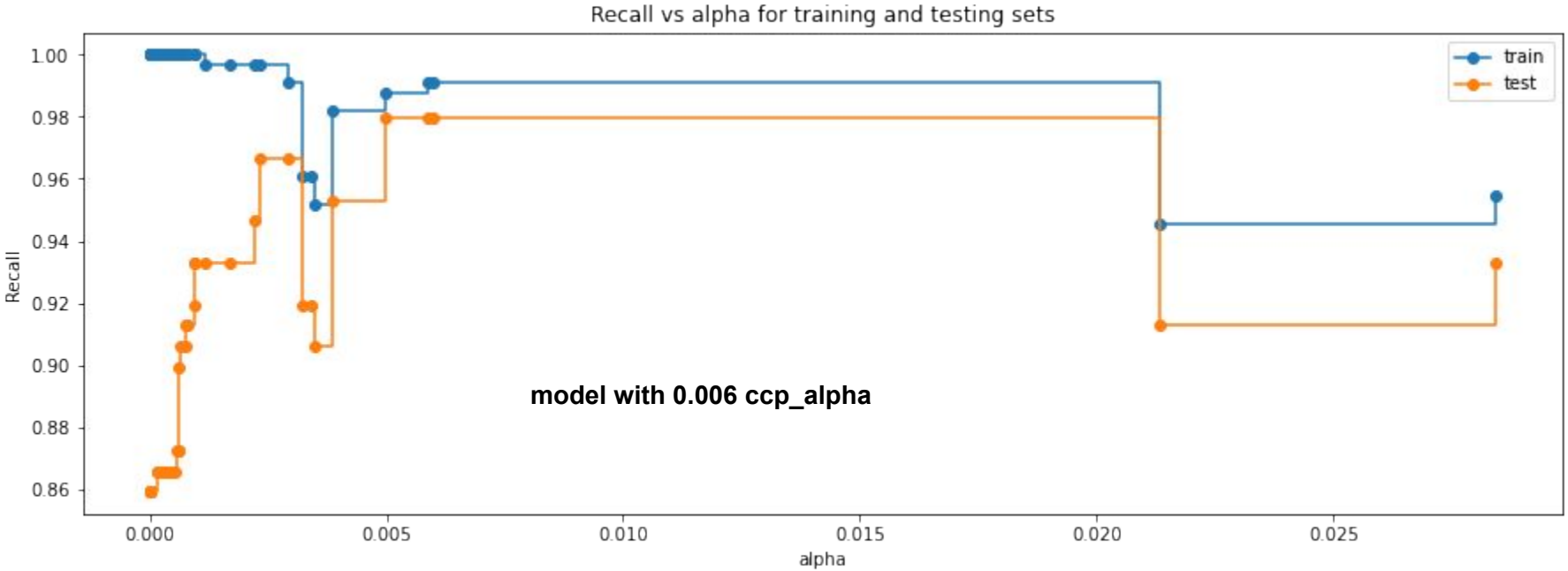
Logistic Regression Model Performance Summary



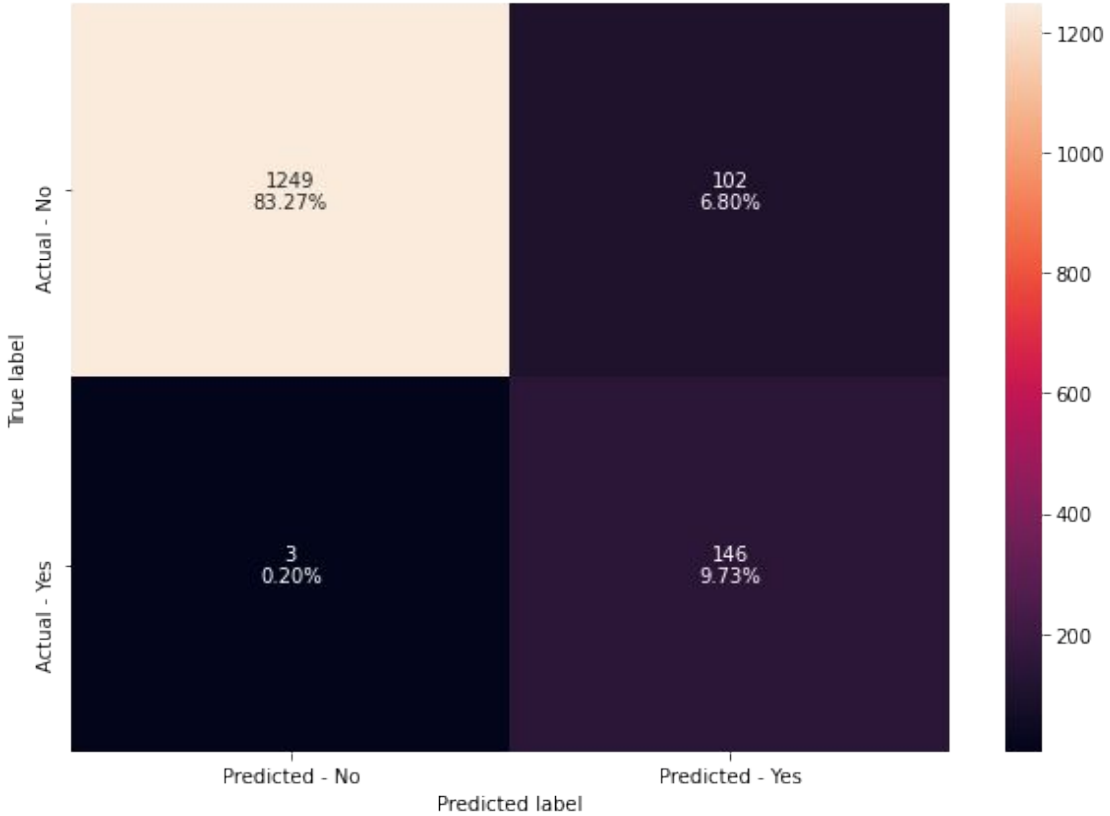
Logistic Regression Model Performance Summary

- Highest performing Logistic Regression model is : * Logistic Regression with feature elimination
- Logistic Regression most important features are :
  - Education
  - CD_Account

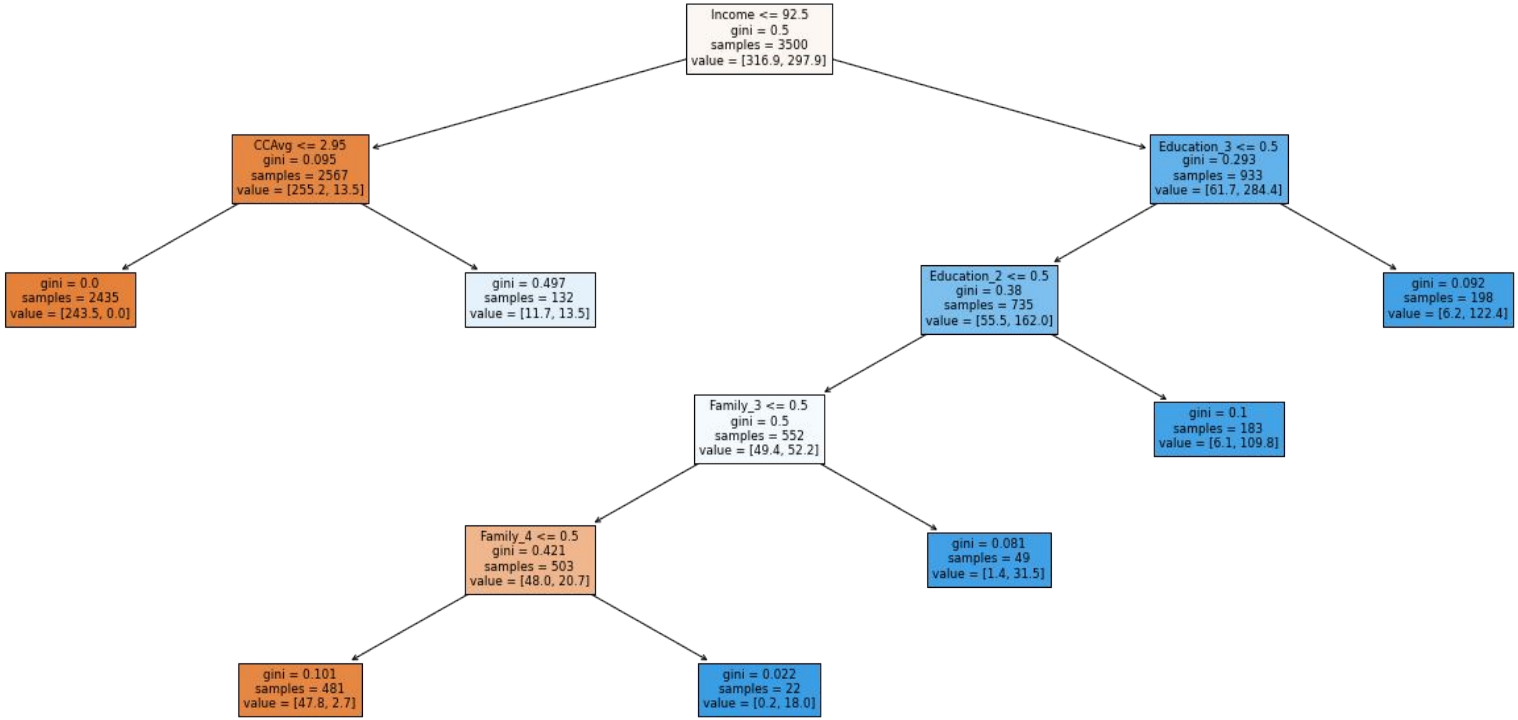
Decision Tree Regression Model Performance Summary



Decision Tree Regression Model Performance Summary



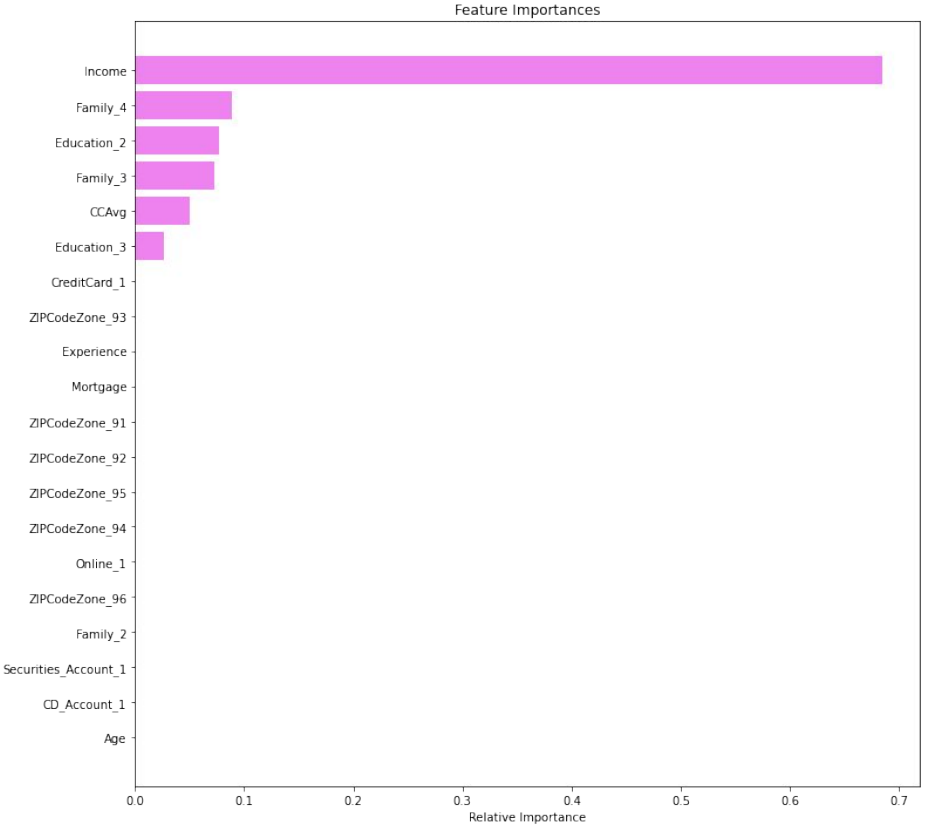
Decision Tree Regression Model Performance Summary



Decision Tree Regression Model Performance Summary

```
|--- Income <= 92.50
|   |--- CCAvg <= 2.95
|   |   |--- weights: [243.50, 0.00] class: 0
|   |--- CCAvg >  2.95
|   |   |--- weights: [11.70, 13.50] class: 1
|--- Income >  92.50
|   |--- Education_3 <= 0.50
|   |   |--- Education_2 <= 0.50
|   |   |   |--- Family_3 <= 0.50
|   |   |   |   |--- Family_4 <= 0.50
|   |   |   |   |   |--- weights: [47.80, 2.70] class: 0
|   |   |   |   |--- Family_4 >  0.50
|   |   |   |   |   |--- weights: [0.20, 18.00] class: 1
|   |   |   |--- Family_3 >  0.50
|   |   |   |   |--- weights: [1.40, 31.50] class: 1
|   |   |--- Education_2 >  0.50
|   |   |   |--- weights: [6.10, 109.80] class: 1
|   |--- Education_3 >  0.50
|   |   |--- weights: [6.20, 122.40] class: 1
```

Decision Tree Regression Model Performance Summary



Decision Tree Model Performance Summary

- Highest performing Desicion Tree model is : Logistic Regression with feature elimination
- Decision Tree most important features are :
  - Income
  - Family_4

Conclusion

- Highest performing Logistic Regression model is :
  - **Logistic Regression with feature elimination**

- Logistic Regression most important features are :
  - Education
  - CD_Account

- Highest performing Decision Tree model is :
  - **Decision Tree with Post-Pruning**

- Decision Tree most important features are :
  - Income
  - Family_4

Recommendations

- It is predicted that customers with higher education and that already have a CD_Account are more likely to convert to Asset Customers.

- It is predicted that Customers with high Income and on average are a family of 4 are more likley to conert to Asset Customers.

greatlearning
Power Ahead

Happy Learning !

Esteban Ordenes

Post Graduate Program in Data Science and Business Analytics

PGP-DSBA-UTA-Dec20-A